Applications of Standards to Harmonize Data between Laboratories and Microarray Platforms of the Toxicogenomics Research Consortium (TRC)

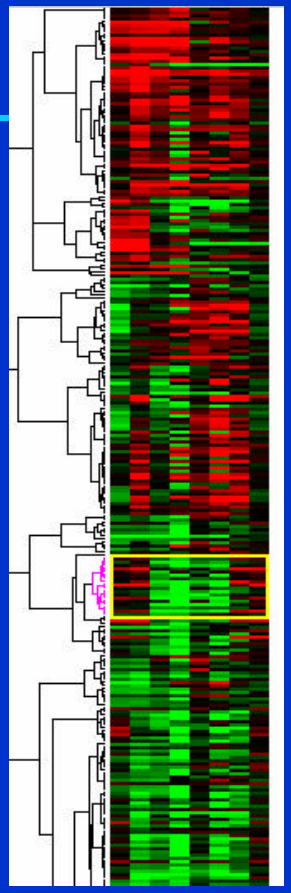
Brenda K. Weis
Extramural Research Coordinator

TRC
Toxicogenomics
Research
Consortium

NCT
National Center
for Toxicogenomics

NIEHS
The National Institute of Environmental Health Sciences
The National Institutes of Health

Toxicogenomics Research Consortium



Main Goals:

1) enhancement of research in the broad area of environmental stress responses using microarray gene expression profiling;

2) provide leadership in toxicogenomics by developing standards and practices for analysis of gene expression data across multiple platforms and address intra- and inter-laboratory variation;

3) development of a robust relational database which combines toxicological end points with changes in gene expression profiles (CEBS);

4) improve public health through better risk detection, earlier intervention in disease processes.

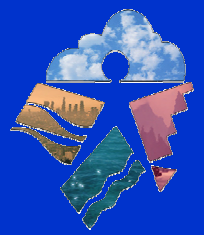
Toxicogenomics Research Consortium

Consists of :

1) ***Cooperative Research Members*** Five academic centers and the NIEHS Microarray Center

2) ***Resource contracts*** in microarray (Paradigm Genetics) and informatics (SAIC)

3) ***Extramural staff:*** Bill Suk (*Program Administrator*), Brenda Weis (*Extramural Toxicogenomics Research Coordinator*), Ben Van Houten (*Science Advisor*), Mike Humble (*Program Analysis*), Jackie Russell (*Grants Management Specialist*)



TRC Cooperative Research Members

- Duke (Schwartz) - developmental defects using zebra fish, innate immunity using LPS (mouse & human), metal toxicity in *C.elegans*
- FHCRC/UW (Zarbl) - transgenic rodents and human cell lines exposed to neurodevelopment or liver toxicants.
- MIT (Sampson) – inter-species (*E.coli*, yeast, WT & KO mice, hepatocytes, bioreactor) responses to alkylating agents
- OHSU (Spencer) - cell-specific injury in the CNS and mechanisms of action of neurotoxicants
- UNC (Kaufmann) - susceptibility factors in genomic response to toxicants (liver, mammary epithelial, nuclear receptor mediated toxicity, alkylating agents)
- NIEHS NMG (Paules) – gene expression profiling of exposure to environmental toxicants

TRC - Components of Support

Core support

- RNA extraction
- Microarray analysis
- Bioinformatics

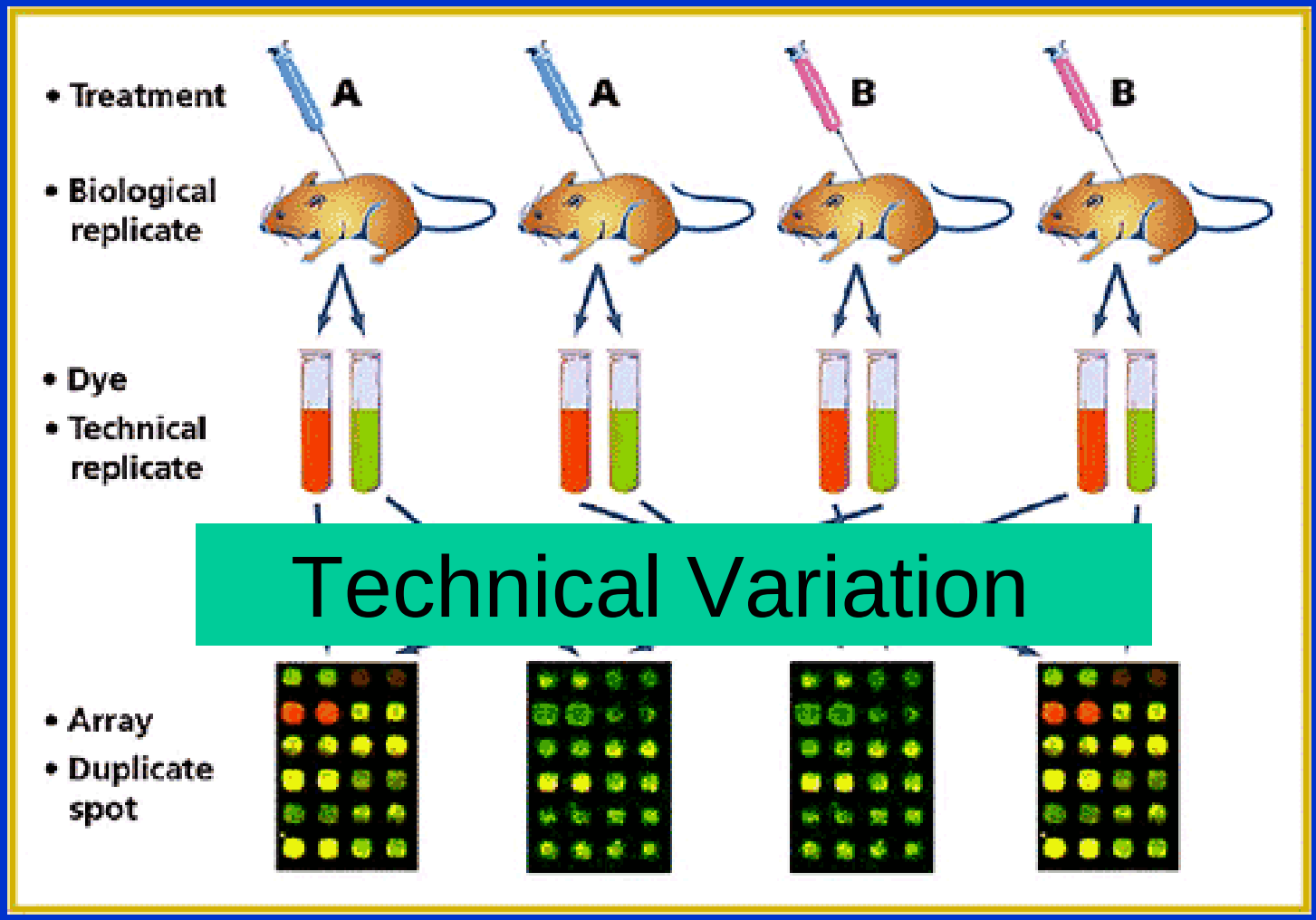
Toxicology experiments

- Cross platform comparisons
- Toxicant signatures
- Trans-species comparisons
- Dose and time effects

Basic research using gene expression profiling

development, cell cycle regulation, signal transduction, metabolism of xenobiotics, cellular responses to injury or stress responses.

Layers of variation in a microarray experiment



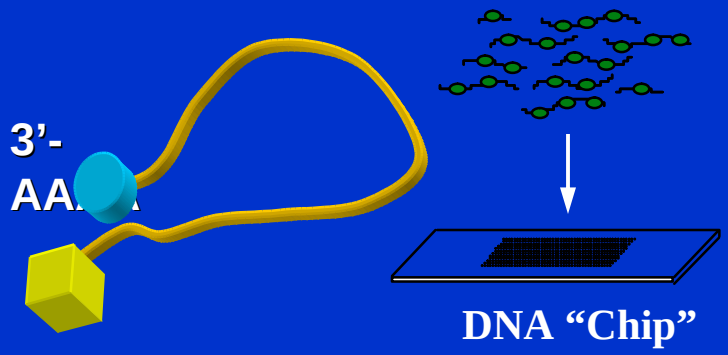
Gary A. Churchill. 2002. Fundamentals of experimental design for cDNA microarrays. Nature Genetics 32, 490 – 495

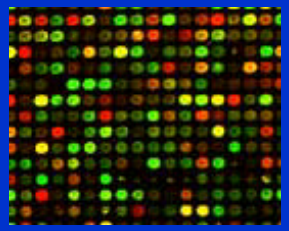
Sources of Variation

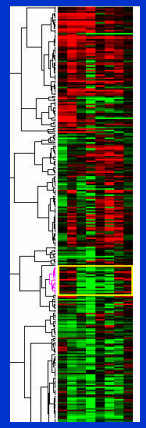
Technical

RNA extraction → Labeling/ Hybridization → Scanning → Analysis

(Bioinformatics)

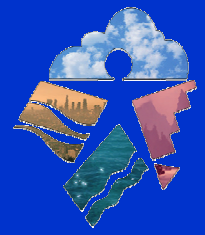






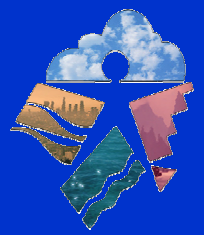
Building a Consortium Project: Where to Start?

Create a “common language” (standardize) for gene expression experiments to generate high quality data and compare and compile data across multiple microarray platforms and laboratories



Importance of Standardization

- Many sources of variation in microarray experiments and application of bioinformatics tools
- Impact of variation on data interpretation unknown
- No standard protocols for the field – have MGED standards, independent source variation unknown
- Currently very difficult to impossible to compare/compile gene expression data across microarray platforms (centers/investigators)
- Needed to consolidate gene expression data in a centralized knowledge database (CEBS)
- Lay foundation for experiments of molecular responses to environmental stressors and risk assessment



Standardization Experiments within the Toxicogenomics Research Consortium

RNA Extraction → Labeling/ Hybridization → Scanning → Analysis (Bioinformatics)

Experiment 1: Determine variation in RNA labeling and hybridization and harmonize protocols across CRMs

Experiment 2: Determine variation in data analysis (bioinformatics) across CRMs

Experiment 3: Determine variation in RNA extraction (toxicant-challenged vs. unchallenged)

Experiment 4: Determine sources of variation in animal husbandry (toxicant-challenged animals/tissue vs. non)

Experiment 1 & 2: Platform Standardization and Analysis

Stnd RNA1 vs Stnd RNA2

Liver vs 5-tissue pool (NIEHS supplied)

Stnd mouse chip (Duke)

17 K oligo chip

6 Cooperative Research Members

- CRM supplied standard RNAs
- CRM supplied standard mouse chip and selects a resident chip
- CRM uses resident protocols for RNA labeling, hyb, scanning, gene annotation, raw data analysis
- Each CRM conducts data analysis for "common genes" on standard and resident chips

Data warehouse (OHSU)

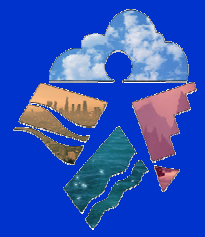
Meta-Analysis: Compare genes on stnd chip across CRMs (Duke-lead)

CRMs analyze same raw data set (**Experiment 2**)

Experiment 3

Design for Experiment 1

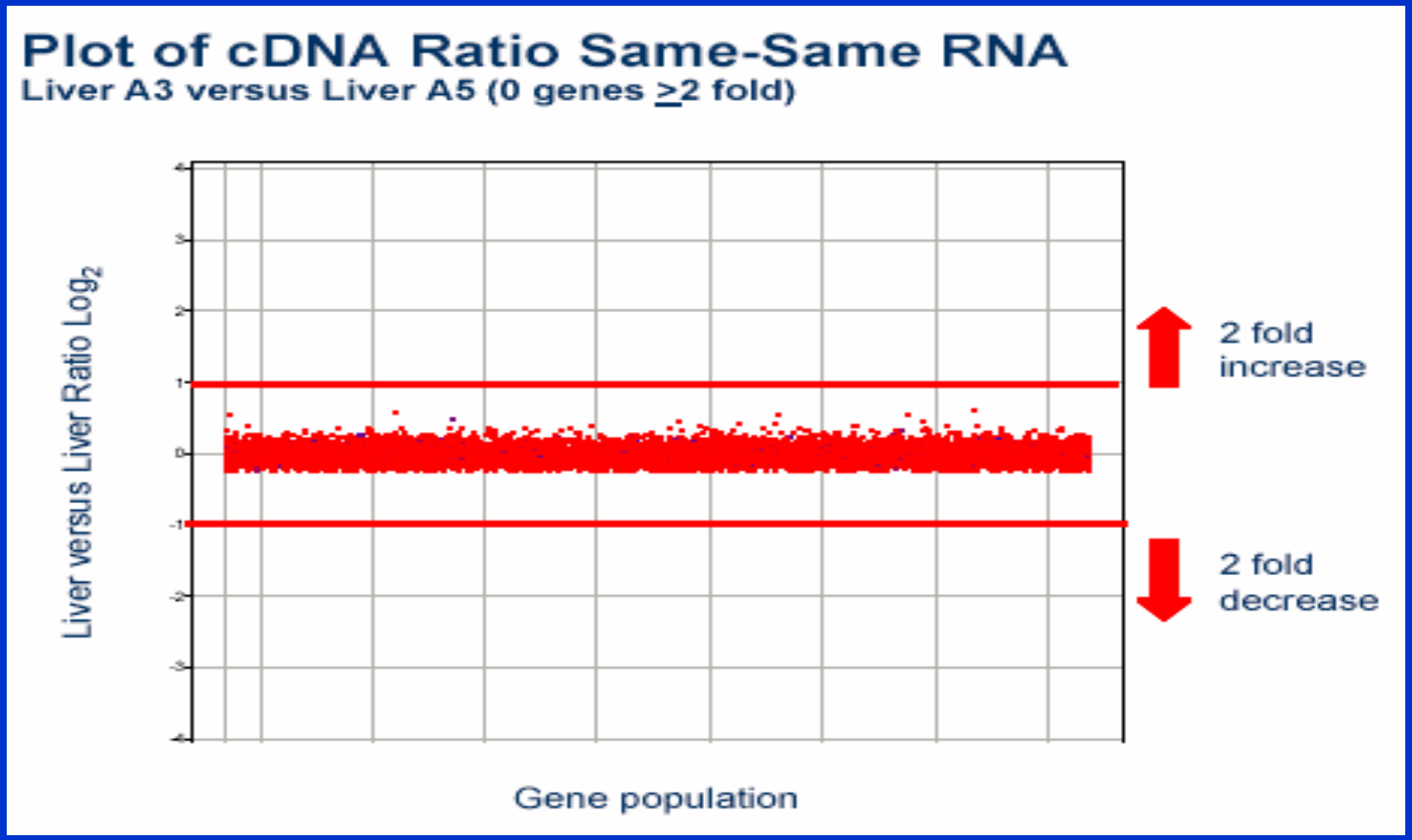
1. Tissue extraction: liver vs. 5-tissue pool from 22 C57 black, adult male mice (NIEHS NTP/OHSU)
2. Standard RNAs: liver, 5-tissue pool (NIEHS)
3. Chips
   - Standard = 18K mouse oligo (Duke)
   - Resident = cDNA, oligo, Affymetrix, Agilent
4. Hybridizations (4): liver vs. liver; liver vs. pooled with fluor dye flips
5. Data quality: Arabidopsis10-gene set in standard RNAs and on chips *(Wang et al, 2002)*
6. OHSU Data Warehouse: Web hosting and data sharing, MIAME sheet for experimental details, gene annotations, resident analysis and stats tools



Consortium Microarray Platforms

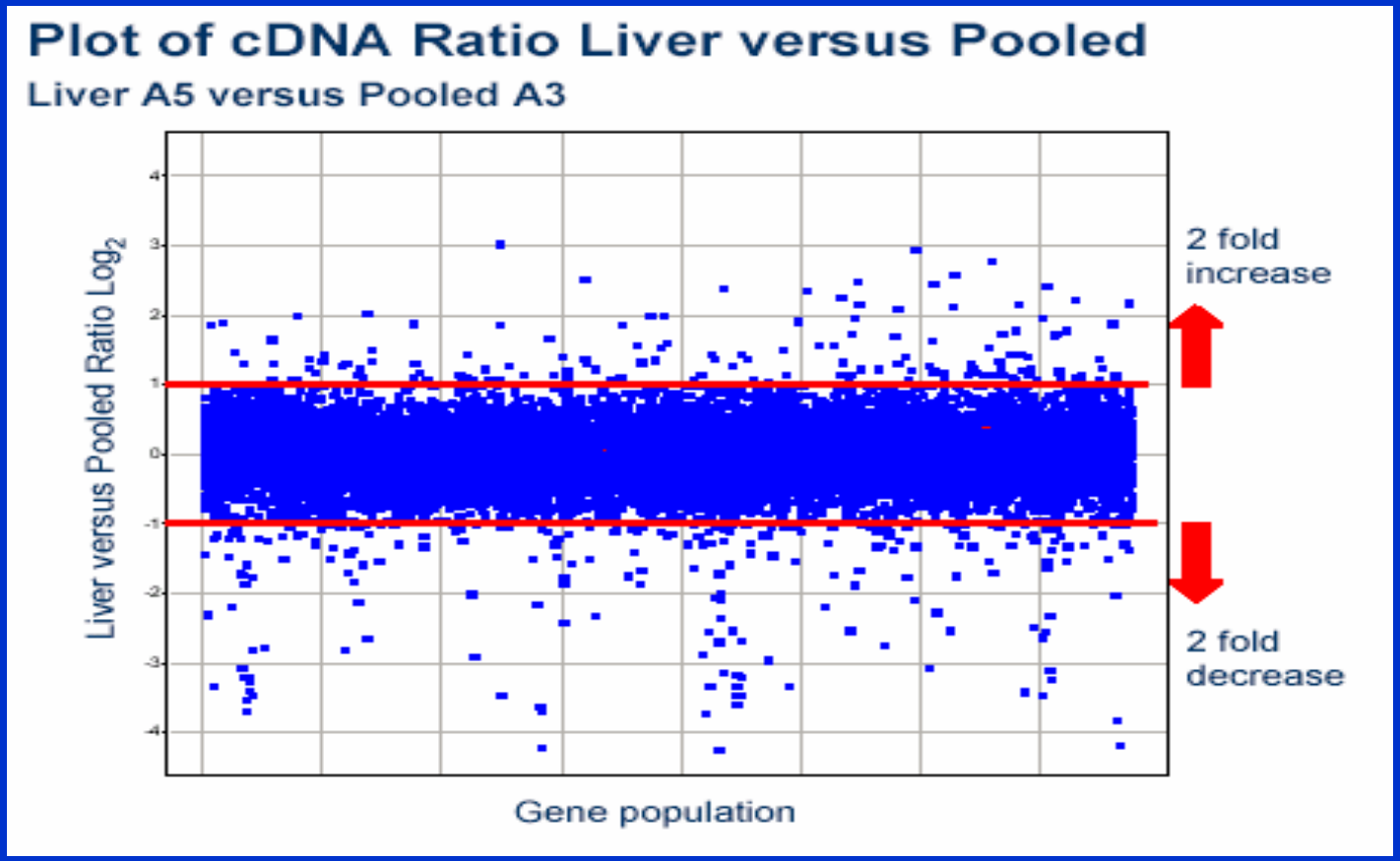
Center	cDNA	Affy	Agilent	Oligo	Oligo (Duke)
FHCRC/UWA	X	X	X		X
MIT	X	X	X		X
Duke	X				X
UNC				X	X
NIEHS (Paradigm)			X X	X	X X
OHSU	X				X

Standard Liver vs. Liver RNA – Duke Chip



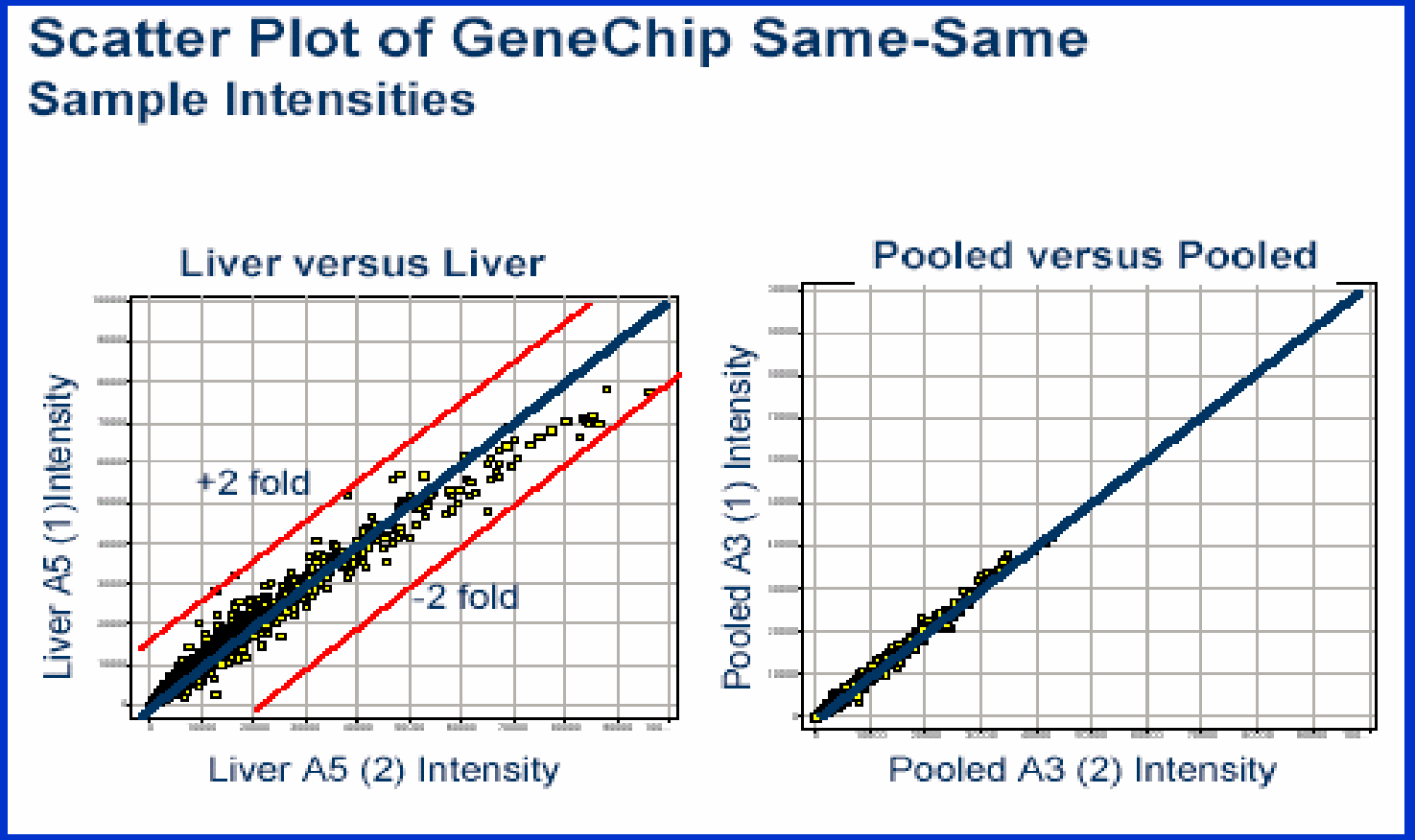
Rebecca Fry, 2003

Standard Liver vs. Pooled RNA – Duke Chip



Rebecca Fry, 2003

Standard RNAs with Affymetrix Chip



Rebecca Fry, 2003

Data Analysis

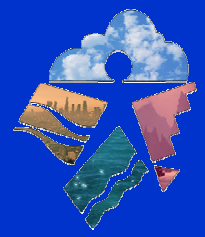
Consortium-wide publications (spring 2003)

Meta-analysis: reliability, reproducibility, quality

- *Common genes* across platforms (3K)
- *Genes on standard chips* across centers (17K)

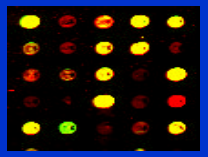
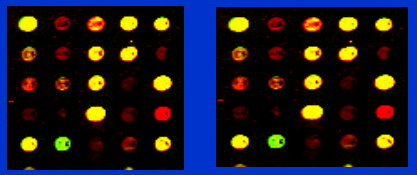
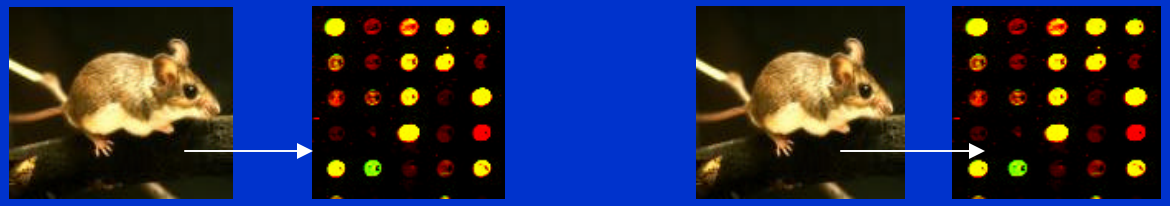
Sources of Technical Variation (n=15, 2003)

- Labeling: direct vs. indirect
- Background correction
- Image analysis/data processing of raw images
- Normalization
- Probe performance (same genes)



"Decreasing correlations reflect the cumulative contributions of multiple sources of variation"

G. Churchill, Nature Genetics 32: 490-495 (2003)

	Correlation
Replicate spots, single microarray 	95%
Same target sample, divided 2 different microarrays 	60-80%
Samples from 2 different inbred mice/same lab 	>30%
Same experiment in different labs	≤30%

Preliminary Trends

Good correlation *within* Centers:

- Standard chip (4 hybridizations)
- Standard chip vs. resident arrays (oligo, cDNA) for 1405 genes (unique Unigene IDs)
- Within and between commercial arrays (Affy, Agilent) - small sample

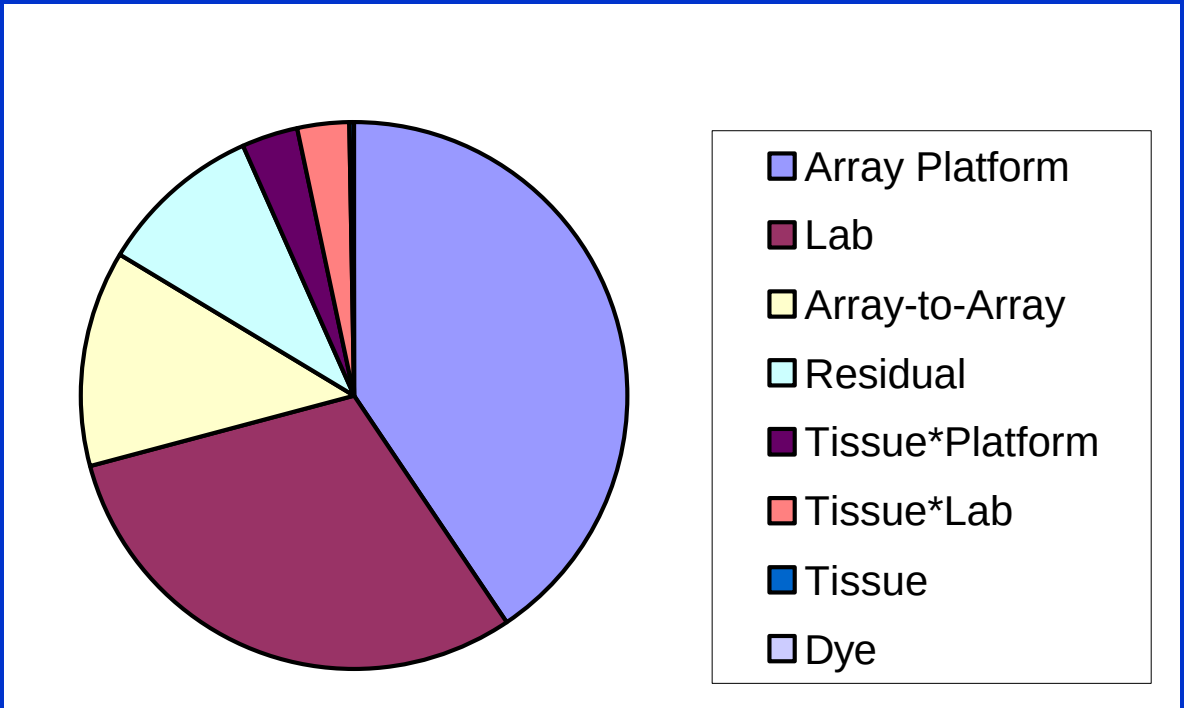
Less and variable correlation *across* Centers:

- Standard chip
- Standard chip vs. resident arrays (oligo, cDNA)
- Resident arrays vs. commercial arrays (Affy, Agilent) for 1100 common elements – small sample

Analysis by UNC/SAS (Deng, Rusyn, Perou and Wolfinger)

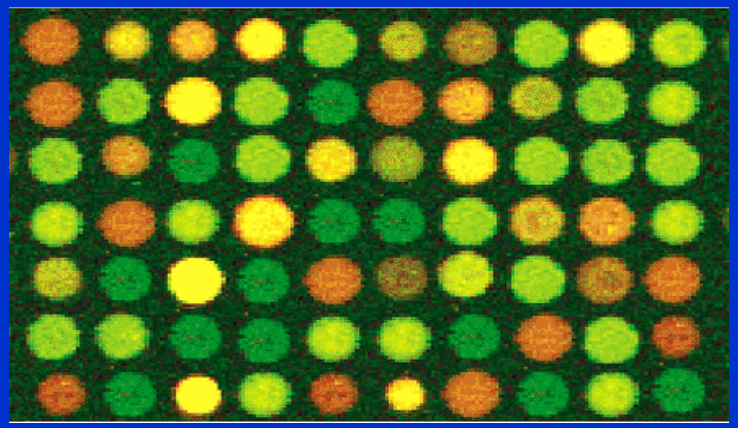
(Wolfinger et al., 2001, *J Computational Biology* 8(6), 625-637)

Sources of Variation – Preliminary Trends



Issues: Gene annotation, preprocessing (background subtraction, signal:noise ratios), flagged data, database issues (need to verify data)

A good prognostic signature



"The Netherlands Cancer Institute in Amsterdam is to become the first institution in the world to use microarray techniques for the routine prognostic screening of cancer patients. Aiming for a June 2003 start date, the center will use a panoply of 70 genes to assess the tumor profile of breast cancer patients and to determine which women will receive adjuvant treatment after surgery."

"…the burgeoning data from the various microarray studies will eventually become standardized…But for now..it makes perfectly good sense."

Standards for Gene Expression Expts

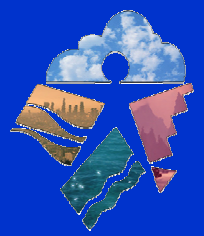
Standard RNAs: single tissue, mixed tissue – reliable, reproducible, stable over time

In vitro transcripts (quality control genes):

- spiked into RNA sample
- serial dilutions of probe randomly on array, cover dynamic range
- correlate spot intensity with transcript dilution

Predictive gene list that can hold up across platforms (species-specific, universal)

Gene annotation: accession number/Unigene cluster, sequence info, commercial arrays



Establishing standards for gene expression

Discussion Points:

- Gene annotation: spend resources on annotating arrays or use arrays with known sequence annotation
- Emerging technologies: RT-PCR as validation step, platform of choice → standards
- Quality control: species-specific or universal standards (Arabidopsis)
- Standards specific for purpose: discovery science, toxicity testing, clinical diagnostic
- Approach: Phased, reference database